

離散数理工学 第9回

離散確率論：確率的離散システムの解析 (発展)

岡本 吉央
okamotoy@uec.ac.jp

電気通信大学

2023年12月12日

最終更新：2023年12月1日 14:17

目次

- ① ランダム・グラフ
- ② ランダム・グラフの閾値現象
- ③ 今日のまとめ

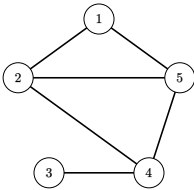
エルデシュ・レニイのランダム・グラフ：例

グラフの集合上の確率分布として、**エルデシュとレニイのランダム・グラフ** が最も古典的で有名

基本的な考え方：次の2つをあらかじめ決めておく

- ▶ n : 頂点数
- ▶ p : 辺確率

$n = 5, p = 1/2$ のときの例：このグラフが得られる確率は $(\frac{1}{2})^{10}$



エルデシュ・レニイのランダム・グラフ：辺確率の独立性

復習：互いに独立であること

「複数の異なる2頂点組 $\{u_1, v_1\}, \dots, \{u_k, v_k\}$ (任意) に対して、辺確率は互いに独立」とは？

任意の $k \in \mathbb{N}$ と、
 複数の異なる任意の2頂点組 $\{u_1, v_1\}, \dots, \{u_k, v_k\}$ に対して

$$\Pr(\{u_1, v_1\} \text{ が辺である, かつ, } \dots, \text{ かつ, } \{u_k, v_k\} \text{ が辺である})$$

$$= \Pr(\{u_1, v_1\} \text{ が辺である}) \cdots \Pr(\{u_k, v_k\} \text{ が辺である})$$

$$= p^k$$

今日の目標

確率的に定義されるグラフの性質を調べられるようになる

- ▶ ランダム・グラフの定義と性質
- ▶ ランダム・グラフの閾値現象 (相転移)

ランダム・グラフ

ランダム・グラフとは？

「確率的に生成されるグラフ」で、次のいずれかを指す

- ▶ 確率的に生成された1つのグラフ
- ▶ グラフの集合上の確率分布

この講義では、「グラフの集合上の確率分布」を考える

エルデシュ・レニイのランダム・グラフ

グラフの集合上の確率分布として、**エルデシュとレニイのランダム・グラフ** が最も古典的で有名

定義：エルデシュとレニイのランダム・グラフ $\mathbb{G}(n, p)$

- ▶ 実数 $p \in (0, 1)$ と正整数 n を考える
- ▶ $\mathbb{G}(n, p)$ は、頂点数 n の無向グラフ全体上の確率分布で、次を満たす
 - ▶ 各2頂点組 $\{u, v\}$ に対して、

$$\Pr(\{u, v\} \text{ が辺である}) = p \quad (\text{辺確率})$$

- ▶ 複数の異なる2頂点組 $\{u_1, v_1\}, \dots, \{u_k, v_k\}$ (任意) に対して、辺確率は互いに独立

以後、 $\mathbb{G}(n, p)$ のグラフの頂点集合を $V = \{1, 2, \dots, n\}$ で表す

エルデシュ、レニイ、ギルバート

エルデシュとレニイのランダム・グラフはエルデシュとレニイの論文 (1959) に登場するがそれと同時期に、ギルバートの論文 (1959) でも登場している



Paul Erdős
(1913–1996)



Alfréd Rényi
(1921–1970)



Edgar Gilbert
(1923–2013)

[https://commons.wikimedia.org/wiki/File:Erdos_budapest_fall_1992_\(cropped\).jpg](https://commons.wikimedia.org/wiki/File:Erdos_budapest_fall_1992_(cropped).jpg)
<http://mathhistory.st-andrews.ac.uk/PictDisplay/Renyi.html>
<https://www.legacy.com/obituaries/dailyrecord/obituary.aspx?n=edgar-nelson-gilbert&pid=165433665&hid=14801>

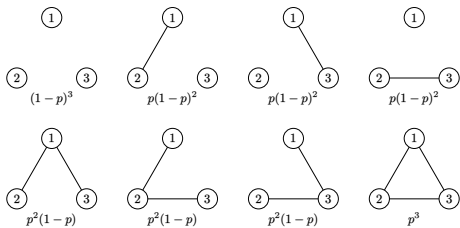
辺確率の独立性から、次がただちに分かる

性質：グラフの生成確率

グラフ G の頂点数が n 、辺数が m のとき、 $\mathbb{G}(n, p)$ において

$$\Pr(G \text{ が生成される}) = p^m (1-p)^{\binom{n}{2}-m}$$

例： $n = 3$ のとき



$G = (V, E)$ を $\mathbb{G}(n, p)$ に従って得られる無向グラフとする

▶ v 以外の任意の頂点 $u \in V - \{v\}$ に対して、

$$\Pr(\{u, v\} \in E) = p$$

▶ 任意の頂点 $u \in V - \{v\}$ に対して、次の標示確率変数 X_u を考える

$$X_u = \begin{cases} 1 & (\{u, v\} \in E), \\ 0 & (\{u, v\} \notin E) \end{cases}$$

▶ このとき、 $\deg(v) = \sum_{u \in V - \{v\}} X_u$

▶ さらに、任意の頂点 $u \in V - \{v\}$ に対して、

$$\mathbb{E}[X_u] = 1 \cdot \Pr(\{u, v\} \in E) + 0 \cdot \Pr(\{u, v\} \notin E) = \Pr(\{u, v\} \in E) = p$$

疑問

$\deg(v)$ は $(n-1)p$ の周辺に集中するの？

チェルノフ上界の技法を用いる

$$\Pr(\deg(v) \geq 2(n-1)p) = \Pr(2^{\deg(v)} \geq 2^{2(n-1)p}) \leq \frac{\mathbb{E}[2^{\deg(v)}]}{2^{2(n-1)p}}$$

ここで、

$$\begin{aligned} \mathbb{E}[2^{\deg(v)}] &= \mathbb{E}\left[2^{\sum_{u \in V - \{v\}} X_u}\right] = \mathbb{E}\left[\prod_{u \in V - \{v\}} 2^{X_u}\right] \\ &= \prod_{u \in V - \{v\}} \mathbb{E}[2^{X_u}] = (2^1 \cdot p + 2^0 \cdot (1-p))^{n-1} = (1+p)^{n-1} \end{aligned}$$

$$\text{したがって、} \Pr(\deg(v) \geq 2(n-1)p) \leq \frac{(1+p)^{n-1}}{2^{2(n-1)p}} = \left(\frac{1+p}{4p}\right)^{n-1}$$

① ランダム・グラフ

② ランダム・グラフの閾値現象

③ 今日のまとめ

頂点 v の **次数** とは、 v に隣接する頂点の数のこと ($\deg(v)$ で表す)

性質： $\mathbb{G}(n, p)$ の次数の期待値

$\mathbb{G}(n, p)$ では、任意の頂点 v に対して、次が成り立つ

$$\mathbb{E}[\deg(v)] = (n-1)p$$

例： $n = 3$ のとき

$$\begin{aligned} \mathbb{E}[\deg(\text{頂点 } 1)] &= 0 \cdot (1-p)^3 + 1 \cdot p(1-p)^2 + 1 \cdot p(1-p)^2 + 0 \cdot p(1-p)^2 \\ &\quad + 2 \cdot p^2(1-p) + 1 \cdot p^2(1-p) + 1 \cdot p^2(1-p) + 2 \cdot p^3 \\ &= 2p(1-p)^2 + 4p^2(1-p) + 2p^3 \\ &= 2p \end{aligned}$$

したがって、

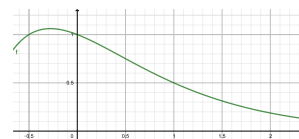
$$\begin{aligned} \mathbb{E}[\deg(v)] &= \mathbb{E}\left[\sum_{u \in V - \{v\}} X_u\right] \\ &= \sum_{u \in V - \{v\}} \mathbb{E}[X_u] \\ &= \sum_{u \in V - \{v\}} p \\ &= (n-1)p \quad \square \end{aligned}$$

疑問

$\deg(v)$ は $(n-1)p$ の周辺に集中するの？

$$\Pr(\deg(v) \geq 2(n-1)p) \leq \left(\frac{1+p}{4p}\right)^{n-1}$$

▶ $p \in (0, 1)$ なので、 $\frac{1}{2} < \frac{1+p}{4p} < 1$



▶ $\therefore 0 < \left(\frac{1+p}{4p}\right)^{n-1} \rightarrow 0 \ (n \rightarrow \infty)$

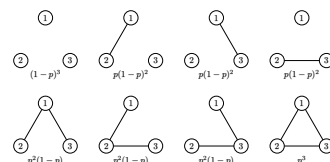
無向グラフの **孤立点**：次数が 0 の頂点のこと

例題

エルデシュとレニイのランダム・グラフ $\mathbb{G}(n, p)$ において、孤立点が存在する確率を計算せよ

$n = 3$ のとき、

$$\begin{aligned} \text{孤立点が存在する確率} &= (1-p)^3 + 3p(1-p)^2 \\ &= (1-p)^2(1+2p) \end{aligned}$$



▶ これを正確に計算するのは難しい

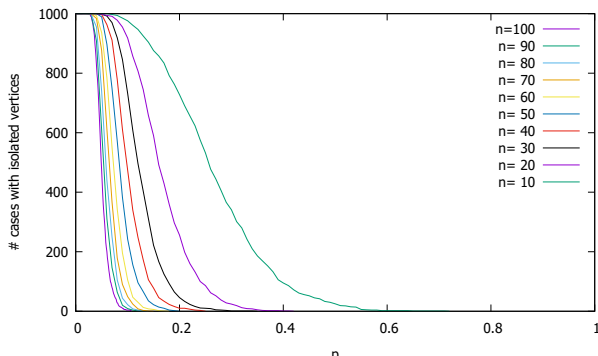
▶ $\therefore n \rightarrow \infty$ のときの漸近的な振る舞いを考察する

注：漸近的な振る舞いの方が (粗い情報なので) 考察しやすい

シミュレーション

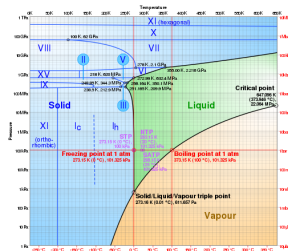
横軸: $p \in [0.0, 1.0]$

縦軸: 1000回の試行中, 孤立点を持つグラフが得られた総数



相転移

相転移 (閾値現象) は 物理学や化学等でも重要な概念



→ ランダム・グラフは物理学者の興味の対象でもある

https://commons.wikimedia.org/wiki/File:Phase_diagram_of_water.svg

ランダム・グラフが孤立点を含む確率: p が大きい場合 (1)

- ▶ 頂点 $v \in V$ に着目
- ▶ G において v が孤立点である確率を計算すると

$$\Pr(v \text{ が孤立点}) = (1 - p)^{n-1}$$

- ▶ 次の確率変数 X_v を考える

$$X_v = \begin{cases} 1 & (v \text{ が孤立点である}), \\ 0 & (v \text{ が孤立点ではない}) \end{cases}$$

- ▶ このとき, G における孤立点の総数 (確率変数) を X とすると

$$X = \sum_{v \in V} X_v, \quad \Pr(G \text{ が孤立点を含む}) = \Pr(X \geq 1)$$

ランダム・グラフが孤立点を含む確率: p が大きい場合 (3)

より一般に, $p \geq (1 + \epsilon) \frac{\ln n}{n-1}$ とする (ただし, $\epsilon > 0$)

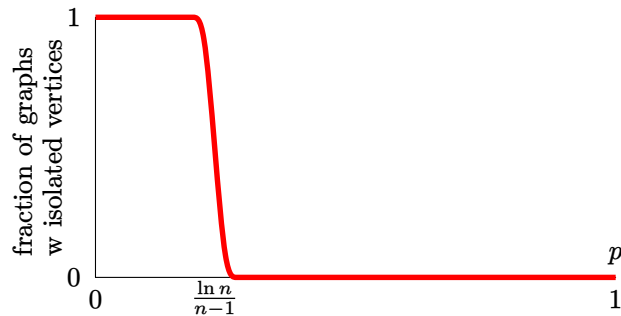
- ▶ このとき, マルコフの不等式より

$$\begin{aligned} \Pr(X \geq 1) &\leq \frac{E[X]}{1} = E[X] = E\left[\sum_{v \in V} X_v\right] = \sum_{v \in V} E[X_v] \\ &= \sum_{v \in V} \Pr(v \text{ は孤立点}) \\ &= \sum_{v \in V} (1 - p)^{n-1} = n(1 - p)^{n-1} \leq ne^{-p(n-1)} \\ &\leq ne^{-(1+\epsilon)\frac{\ln n}{n-1}(n-1)} = ne^{-(1+\epsilon)\ln n} = n \cdot \frac{1}{n^{1+\epsilon}} \\ &= \frac{1}{n^\epsilon} \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

- ▶ つまり, $\Pr(G \text{ が孤立点を含む}) \rightarrow 0 \quad (n \rightarrow \infty)$

理論: 今から証明すること

$n \rightarrow \infty$ のとき



「孤立点を持つ」という性質は 閾値現象 を持つ (相転移 を起こす)

ランダム・グラフが孤立点を含む確率: 先に結論

G は $\mathbb{G}(n, p)$ に従って得られるグラフとする

性質: ランダム・グラフが孤立点を含む確率

$\epsilon \in (0, 1)$ は任意の正実数

- ▶ $p \leq (1 - \epsilon) \frac{\ln n}{n-1}$ のとき

$$\Pr(G \text{ が孤立点を含む}) \rightarrow 1 \quad (n \rightarrow \infty)$$

- ▶ $p \geq (1 + \epsilon) \frac{\ln n}{n-1}$ のとき

$$\Pr(G \text{ が孤立点を含む}) \rightarrow 0 \quad (n \rightarrow \infty)$$

感覚をつかむため, まず $\epsilon = \frac{1}{2}$ のときに証明する

ランダム・グラフが孤立点を含む確率: p が大きい場合 (2)

$p \geq \frac{3}{2} \frac{\ln n}{n-1}$ とする

- ▶ このとき, マルコフの不等式より

$$\begin{aligned} \Pr(X \geq 1) &\leq \frac{E[X]}{1} = E[X] = E\left[\sum_{v \in V} X_v\right] = \sum_{v \in V} E[X_v] \\ &= \sum_{v \in V} \Pr(v \text{ は孤立点}) \\ &= \sum_{v \in V} (1 - p)^{n-1} = n(1 - p)^{n-1} \leq ne^{-p(n-1)} \\ &\leq ne^{-\frac{3}{2}\frac{\ln n}{n-1}(n-1)} = ne^{-\frac{3}{2}\ln n} = n \cdot \frac{1}{n^{3/2}} \\ &= \frac{1}{n^{1/2}} \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

- ▶ つまり, $\Pr(G \text{ が孤立点を含む}) \rightarrow 0 \quad (n \rightarrow \infty)$

ランダム・グラフが孤立点を含む確率: p が小さい場合 (1)

$p \leq (1 - \epsilon) \frac{\ln n}{n-1}$ とする

- ▶ 先ほどと同様に, X_v と X を定義する
- ▶ このとき, $\Pr(G \text{ が孤立点を含まない}) \rightarrow 0$ を示したいが

$$\begin{aligned} \Pr(G \text{ が孤立点を含まない}) &= \Pr(X = 0) \\ &\leq \Pr(X \geq 0) \\ &= 1 \end{aligned}$$

単純にマルコフの不等式が使えないので, 困る...

解決策

チェビシエフの不等式を使う

定理：チェビシェフの不等式

確率変数 X と正実数 $t > 0$ に対して、

$$\Pr(|X - E[X]| \geq t) \leq \frac{E[X^2] - E[X]^2}{t^2}$$

証明：マルコフの不等式より

$$\Pr(|X - E[X]| \geq t) = \Pr(|X - E[X]|^2 \geq t^2) \leq \frac{E[|X - E[X]|^2]}{t^2}$$

ここで、

$$\begin{aligned} E[|X - E[X]|^2] &= E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2] \\ &= E[X^2] - 2E[E[X]X] + E[E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2 \quad \square \end{aligned}$$

$$\begin{aligned} E[X^2] &= E\left[\left(\sum_{v \in V} X_v\right)^2\right] = E\left[\sum_{u \in V} \sum_{v \in V} X_u X_v\right] \\ &= E\left[\sum_{v \in V} X_v^2 + \sum_{u \in V} \sum_{v \in V - \{u\}} X_u X_v\right] \\ &= E\left[\sum_{v \in V} X_v + \sum_{u \in V} \sum_{v \in V - \{u\}} X_u X_v\right] \quad (\because X_v \text{ は } 0 \text{ か } 1) \\ &= E\left[\sum_{v \in V} X_v\right] + E\left[\sum_{u \in V} \sum_{v \in V - \{u\}} X_u X_v\right] \\ &= E[X] + \sum_{u \in V} \sum_{v \in V - \{u\}} E[X_u X_v] \end{aligned}$$

以上の議論をまとめると、 $p \leq \frac{1}{2} \frac{\ln n}{n-1}$ のとき、

$$\begin{aligned} \Pr(G \text{ が孤立点を含まない}) &\leq \frac{E[X^2]}{E[X]^2} - 1 \\ &= \frac{n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}}{(n(1-p)^{n-1})^2} - 1 \\ &= \frac{1}{n(1-p)^{n-1}} + \frac{n(n-1)(1-p)^{2n-3}}{n^2(1-p)^{2n-2}} - 1 \\ &\leq \frac{1}{n(1-p)^{n-1}} + \frac{n^2(1-p)^{2n-3}}{n^2(1-p)^{2n-2}} - 1 \\ &= \frac{1}{n(1-p)^{n-1}} + \frac{1}{1-p} - 1 \\ &\leq \frac{1}{n} e^{p(n-1)} + e^p - 1 \end{aligned}$$

より一般に、 $p \leq (1-\epsilon) \frac{\ln n}{n-1}$ のとき (ただし、 $\epsilon > 0$)、

$$\begin{aligned} \Pr(G \text{ が孤立点を含まない}) &\leq \frac{1}{n} e^{p(n-1)} + e^p - 1 \\ &\leq \frac{1}{n} e^{(1-\epsilon) \frac{\ln n}{n-1} (n-1)} + e^{(1-\epsilon) \frac{\ln n}{n-1}} - 1 \\ &= \frac{1}{n} n^{1-\epsilon} + e^{(1-\epsilon) \frac{\ln n}{n-1}} - 1 \\ &= \frac{1}{n^\epsilon} + e^{(1-\epsilon) \frac{\ln n}{n-1}} - 1 \\ &\rightarrow 0 + e^0 - 1 \quad (n \rightarrow \infty) \\ &= 0 \\ \therefore \Pr(G \text{ が孤立点を含む}) &\rightarrow 1 \quad (n \rightarrow \infty) \end{aligned}$$

$$\begin{aligned} \Pr(G \text{ が孤立点を含まない}) &= \Pr(X = 0) \\ &\leq \Pr(X = 0 \text{ または } X = 2E[X]) \\ &= \Pr(|X - E[X]| = E[X]) \\ &\leq \Pr(|X - E[X]| \geq E[X]) \\ &\leq \frac{E[X^2] - E[X]^2}{E[X]^2} \\ &= \frac{E[X^2]}{E[X]^2} - 1 \end{aligned}$$

まとめ

- ▶ 既に計算済み： $E[X] = n(1-p)^{n-1}$
- ▶ いまから計算： $E[X^2] = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}$

u と v が異なる頂点であるとき

$$\begin{aligned} E[X_u X_v] &= \Pr(X_u X_v = 1) = \Pr(X_u = 1 \text{ かつ } X_v = 1) \\ &= \Pr(u \text{ と } v \text{ がともに孤立点である}) \\ &= (1-p)^{2n-3} \end{aligned}$$

したがって、

$$\sum_{u \in V} \sum_{v \in V - \{u\}} E[X_u X_v] = n(n-1) \cdot (1-p)^{2n-3}$$

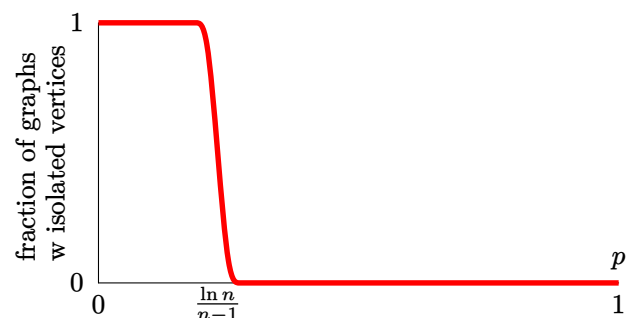
したがって、

$$\begin{aligned} E[X^2] &= E[X] + \sum_{u \in V} \sum_{v \in V - \{u\}} E[X_u X_v] \\ &= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} \end{aligned}$$

以上の議論をまとめると、 $p \leq \frac{1}{2} \frac{\ln n}{n-1}$ のとき、

$$\begin{aligned} \Pr(G \text{ が孤立点を含まない}) &\leq \frac{1}{n} e^{p(n-1)} + e^p - 1 \\ &\leq \frac{1}{n} e^{\frac{1}{2} \frac{\ln n}{n-1} (n-1)} + e^{\frac{1}{2} \frac{\ln n}{n-1}} - 1 \\ &= \frac{1}{n} n^{1/2} + e^{\frac{1}{2} \frac{\ln n}{n-1}} - 1 \\ &= \frac{1}{n^{1/2}} + e^{\frac{1}{2} \frac{\ln n}{n-1}} - 1 \\ &\rightarrow 0 + e^0 - 1 \quad (n \rightarrow \infty) \\ &= 0 \\ \therefore \Pr(G \text{ が孤立点を含む}) &\rightarrow 1 \quad (n \rightarrow \infty) \end{aligned}$$

$n \rightarrow \infty$ のとき



「孤立点を持つ」という性質は **閾値現象** を持つ (**相転移** を起こす)

- ① ランダム・グラフ
- ② ランダム・グラフの閾値現象
- ③ 今日のまとめ

今日の目標

確率的に定義されるグラフの性質を調べられるようになる

- ▶ ランダム・グラフの定義と性質
- ▶ ランダム・グラフの閾値現象 (相転移)